Librarian Symposium Report

Document Control Information

Settings	Value
Document Identifier:	D6.22
Project Title:	ExPaNDS
Work Package:	WP6
Document Author(s):	Isabelle Boscaro-Clarke (Diamond), Kat Roarty (Diamond), Rebecca Stredwick (Diamond)
Document reviewer(s):	Brian Matthews (UKRI), Abigail McBirnie (UKRI)
Responsible Partner:	Diamond
Doc. Issue:	1.0
Dissemination level:	Public
Date:	04/02/2022

Abstract

In the frame of our engagement strategy towards our librarian and data managers stakeholder group, we held a Librarian Symposium in September 2021. This report highlights the discussion points and key insights gained during the webinar.

Licence

ExPaNDS

ExPaNDS Symposium for Librarians and Data Managers

The ExPaNDS Symposium for Librarians and Data Managers took place on 30th September 2021 and was organised by Work Packages 2 (WP2) and 6 (WP6) from the ExPaNDS project. The Symposium focused on the interface between publications and data, asking:

1. How can we bring these together better for the benefit of all?
2. Where are the gaps?
3. What do we need to bridge these gaps?

The agenda covered data licensing, legislation, and linking and relationships – e.g. how different research and data are published and related together to create a traceable, findable environment; and impact and FAIR metrics – e.g. how we go about measuring the effect of collecting metrics and generating the outputs we provide and publishing the experiment. Data is only valuable if it can be used and shared, but both need to be done responsibly.

Dr Brian Matthews (ExPaNDS Lead of WP2: Enabling FAIR Data) and Dr Abigail McBirnie (also ExPaNDS WP2) chaired the sessions. Both work for UK Research and Innovation (UKRI) in the Scientific Computing Department at the Science and Technology Facilities Research Council (STFC). Brian is Group Leader of the Open Data Systems Group, and Abigail is a Senior Research Officer in Open Science.

Introducing ExPaNDS
Brian opened the Symposium by introducing ExPaNDS and its work promoting FAIR (Findable, Accessible, Interoperable and Reusable) data and data services for facilities through enabling FAIR policies, guidelines, tools, and experiments. He then outlined the important role Librarians and Data Managers have in making research FAIR, from curating and maintaining records of science through to communicating these to the research community and recording the citation and impact of science.

Data licensing and legislation
The first sessions on Data Licensing and Legislation were led by Data Lawyer - Dr Lauro Fava, Associate at the law firm, Pinsent Masons, and Dr Rebecca Grant, Head of Data and Software Publishing at F1000 Research. Presentations covered Data Law and different data licences (liberal to restrictive).

Discussions focused on the challenges caused by the patchwork of legal protection and the variation between countries. Contracts are therefore essential to settle any uncertainties when sharing data. There should be more prominence of the roles of Librarians/Data Managers in providing guidelines, support and information to researchers. Awareness of repository licence policy and journal editorial discretion is useful. It is possible to have two different licences on the same data set/publish in multiple repositories. The user can choose what set of conditions would be best. Moving from a more restrictive licence to a more open one is not possible; the past licence will still apply.

Key points included:
- Data Law is not about ownership but rather about rights;
- Data licence is a form of contract;

- The specific expression of the idea is subject to copyright but not the idea itself;
- Contract is king;
- Contract terminology needs to be specific;
- Challenges include a lack of harmonisation around data terminology;
- Data citation is not always mandatory in publisher policy;
- There is a strong author preference for more restrictive licences;
- Force 11 Joint Declaration of Data Citation Principles provides information on Data Attribution.

Questions for further consideration:
- Should or do publishers recommend certain repositories?
- Do researchers understand the implications of different licences and of using one over the other?
- How do we direct authors to repositories which supply acceptable licences?
- Does data licensing work without mandatory data citation?
- Do we need closer collaboration between publishers and data repositories to encourage good practice?

Linking and Inter-relationships using Persistent Identifiers (PIDs) and the PUMA project

Dr Vasily Bunakov, a Senior Researcher in Open Data Systems from UKRI/STFC also working within ExPaNDS WP2, looked at linking and Inter-relationships using Persistent Identifiers (PIDs). This was followed by a demonstration of the Publication and User experiment Metadata Analyser (PUMA) project by Renaud Duyme, a Developer at The European Synchrotron (ESRF). PUMA is a tool developed to assist in the effective identification of the relationship of data to proposal and experiment, through to data analysis and publication.

Vasily explained the different types of PIDs and clarified common misconceptions. Discussion centred on how to embed PIDs in our systems, how current PIDs can be better used, and how best to use PIDs for making connections and exploring trends. Advocacy is needed to demonstrate PID impact and change the culture of underuse of PIDs. The key is to adopt best practices and to explore how to instigate these. It is therefore essential to allocate time and financial resources for this.

Renaud showed how PUMA allows powerful reporting on beamlines, techniques, science topics, member country activities, and metadata, and supports strategic planning. Automated searching and matching reveals who else is using and participating in the data. Full text searches find affiliations and access repositories.

Key insights covered:
- PIDs can be used for data retrieval and insights on research;
- It is much easier to use PIDs from the start of the research process than later;
- PID graphs to link PIDs offer huge benefits if we start using them well;
- The challenge in PUMA is linking proposals, experiment reports and publications, and categorising consistently;
- More information on PIDs can be found in Research Data Alliance groups; the Persistent Identifier policy for the European Open Science Cloud; and the PID Forum, amongst others.

Impact and FAIR metrics

UKRI/STFC's Mark Thorley introduced access to research data from public funding – the OECD recommendation. Mark is the coordinator of the StR-ESFRI2 project and is involved in developing the OECD recommendations.

The OECD covers 38 member countries including all ExPaNDS partners' countries. OECD Recommendations on Research Data focus on seven policy areas and were adopted in 2006 and amended in 2021. They range from Data Governance for Trust and Technical Standards and Practices to Human Capital and International Cooperation for Access to Research Data, with the aim of supporting best practices.

The following points were discussed:
- Scientific data is viewed as a public good and should be provided without discrimination;
- Policy framework is worked on together; member states report back on their actions;
- It is explicit that data should be FAIR;
- Technical standards and practices should be FAIR, with open licences and standards;
- Recommendations are extended to other research relevant digital objects;
- Create a trusted environment and FAIR culture that allows researchers to be confident about when to say no and when to say yes to data sharing;
- Developing a skills base is key, including career paths to attract and retain data scientists and dedicated Data Stewards;
- Research data is part of the national policy driver;
- It is vital to work with communities to develop targets and progress reporting.

The next session was led by Dr Robert Huber, from Universität Bremen in Germany. Robert is a Marine biologist responsible for scientific data management. He is a lead developer of the F-UJI Automated FAIR Data Assessment Tool. He showed how FAIR Data Assessment Pilots in FAIRsFAIR have developed assessment metrics and practical tools to meet recommendations by the European Commission's Turning FAIR into Reality report (2018)[1]. These tools support a very large and diverse community. It is important to ensure processes to generate data produce a FAIR result.

Key insights from the session:
- FAIR data assessments pilots setting and evaluating assessment metrics;
- These apply to principles and to practical tests;
- Assessment scenarios 1&2 – FAIR Aware – raise awareness before data is deposited;
- Assessment scenarios 6&7 – the F-UJI Tool - is a FAIR assessment of already published data FAIRsFAIR data metrics align with the RDA and WDS metrics[2];
- F-UJI auto assesses metadata from PIDs, gives a rating on FAIRness, and uses domain agnostic metadata standards;
- Datasets can be tested via this link, which also produces sharable reports: F-UJI.

---

[1] https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283 accessed on 16.11.2021

[2] https://www.rd-alliance.org/groups/assessment-data-fitness-use accessed on 16.11.2021

Discussion focused on using domain specific formats and vocabularies and how do deal with more specialised domains. The recommendation is to always use domain agnostic and domain specific metadata - both are helpful in metadata frameworks. Future work is needed to handle more domain specific data.

ublishing the experiment
Dr Oliver Knodel, data scientist, post-doc and research software engineer at the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) in Germany, introduced the motivation for publishing experiments and also demonstrated the PaN Training Catalogue, under development in ExPaNDS. In addition, tools for licences for software like GitHUB were introduced.

The presentation outlined the challenges that researchers face when organising data, choosing and accessing repositories, working with copyright, licensing, and costs of sharing data, and choosing formats and open standards that enable interoperability. Establishing a FAIR data policy framework at each facility is vital to provide guidance and make the process easier.

The PaN Training Catalogue was developed to provide an easy-to-use environment for references to external resources and training workflows, including on how to publish and use FAIR data. It is possible to describe relationships between methods and data in an overall training workflow, bringing everything together. Steps in experiments can be made visible to other scientists in the interests of FAIR data.

Discussions centred on the challenge of balancing openness with concerns about competition and whether scientists are prepared to show each step of their research (including aborted/failed attempts).

Role of a Librarian/Data Managers within FAIR research lifecycle?
In the final session, Dr Brian Matthews explored the role of Librarian/Data Managers as guides, supporters and quality controllers who make sure everything is provided to help researchers along the research journey, from training researchers in FAIR data good practice (such as data handling), to guiding in the use of tools, and reporting on the impact of experiments. There is a lot of good work happening. Data Managers and Librarians already meet in several arenas and through FAIR data collaborations.

Brian asked: Are Librarian/Data Manager roles evolving in these directions and what kind of balance should there be in delivery?

In the discussion, the following recommendations were made:
- Sufficient resources are needed to enable Librarians/Data managers in these tasks;
- Engage with user communities through in-depth presentations on specific questions like licences;
- Tailor to the needs of each facility. Each is at a different stage of the FAIR journey;
- More in-depth meetings are very useful and so is feedback.

If you'd like to view the recording of this workshop, please visit the ExPaNDS website.